



**Fakultät für  
Informatik und Mathematik**

# **Bachelorarbeit**

über das Thema

**Entwurf und Implementierung eines Chrome Plugins zur  
automatischen Anreicherung von Webseiten mit kulturellen  
Inhalten**

**Autor:** Mathias Möller  
moellerm@fim.uni-passau.de

**Prüfer:** Prof. Granitzer

**Abgabedatum:** 24.08.2015

## I Kurzfassung

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

## Abstract

Das ganze auf Englisch.

## II Inhaltsverzeichnis

<b>I</b>	<b>Kurzfassung</b>	<b>I</b>
<b>II</b>	<b>Inhaltsverzeichnis</b>	<b>II</b>
<b>III</b>	<b>Abbildungsverzeichnis</b>	<b>IV</b>
<b>IV</b>	<b>Tabellenverzeichnis</b>	<b>V</b>
<b>V</b>	<b>Listing-Verzeichnis</b>	<b>V</b>
<b>VI</b>	<b>Abkürzungsverzeichnis</b>	<b>VI</b>
<b>1</b>	<b>Einleitung und Motivation</b>	<b>1</b>
1.1	Unterschiede Web Augmentation, Web Personalization, Web Customization	2
1.2	Vormarsch von JITIR/Web Augmentation . . . . .	2
1.3	Verbesserung des EEXCESS Plugins / Unterschiede zum EEXCESS Plugin	2
<b>2</b>	<b>Related Work</b>	<b>3</b>
2.1	Unterschiede und Gemeinsamkeiten zu JITIR-Agents . . . . .	3
2.1.1	Proaktiv vs. User Interaction . . . . .	3
2.1.2	Informationen darstellen in „Nonintrusive Manner“ . . . . .	3
2.1.3	Awareness of user's local context . . . . .	3
2.2	Vergleich mit existierenden JITIR Agents . . . . .	3
2.2.1	Remembrance Agent (in EMACS Editor) . . . . .	3
2.2.2	Margin? Web Plugin ähnlich wie EEXCESS . . . . .	3
2.2.3	Watson . . . . .	3
2.2.4	EEXCESS? . . . . .	3
2.3	Text Retrieval Algorithmen . . . . .	3
2.4	Unterschiede und Gemeinsamkeiten zu „Automatic help systems“ (z.B. Microsoft Office Assistant - ı Domain spezifisch) . . . . .	3
<b>3</b>	<b>Konzept</b>	<b>3</b>
3.1	Warum kein Proactiver JITIR-Agent? . . . . .	3
3.2	Anzeige der Ergebnisse . . . . .	4
3.3	Erklärung der Such-Anfragen Generierung . . . . .	4
3.4	Anpassen der Suchanfrage durch den Nutzer . . . . .	4
3.5	Verbesserung der Suchanfrage z.B. durch maschinelles Lernen . . . . .	4
<b>4</b>	<b>Implementierung</b>	<b>4</b>
4.1	Verwendung von AngularJS für alle Komponenten des Plugins . . . . .	4
4.2	Bau der GUIs . . . . .	4
4.3	Einbindung der REST-Services . . . . .	4
<b>5</b>	<b>Evaluierung?</b>	<b>4</b>
<b>6</b>	<b>Future Work</b>	<b>4</b>

6.1	Alternative Algorithmen zur Textanalyse . . . . .	4
6.2	Implementierung einer automatischen Suchanfragen-Verbesserung durch ma- schinelles Lernen . . . . .	4
6.3	Verbesserung der Ergebniss-Güte . . . . .	4
6.4	Anpassung der Anwendung auf mobile Nutzung . . . . .	5
<b>7</b>	<b>Conclusion</b>	<b>5</b>
<b>8</b>	<b>Quellenverzeichnis</b>	<b>6</b>
	<b>Anhang</b>	<b>I</b>

### **III Abbildungsverzeichnis**

## **IV Tabellenverzeichnis**

## **V Listing-Verzeichnis**

## **VI Abkürzungsverzeichnis**

# 1 Einleitung und Motivation

Suchmaschinen gehören heutzutage zu den meistbesuchten Seiten im Internet. Google<sup>1</sup> ist mit 2.000.000.000.000 Suchanfragen pro Jahr (TODO: Quelle) nicht nur die beliebteste Suchmaschine, sondern auch die meist besuchte Seite im Internet. Auch die chinesische Suchmaschine Baidu<sup>2</sup> ist in den oberen fünf Plätzen vertreten [AI]. Doch der Funktionsumfang dieser Seiten ist begrenzt. Sie können den Kontext des Benutzers in ihren Queries nicht miteinbeziehen. Weiterhin verlangen sie vom Nutzer seine gegenwärtige Arbeit einzustellen, die Webseite der Suchmaschine aufzurufen und eine Suchanfrage zu formulieren. So wird die Konzentration auf den eigentlichen Arbeitsschritt gestört. Im Schnitt sind nur 67% der Suchanfragen an Google erfolgreich (TODO: Quelle). Das heißt in circa einem Drittel der Fälle bleibt der Aufruf der Suchmaschine ohne die erwarteten Erfolge. Zu der verstrichenen Zeit kommt auch noch die Zeit dazu, die der Nutzer braucht um sich seine Arbeit wieder ins Kurzzeitgedächtnis zu holen. Um diesen Problemen entgegen zu wirken wurden neue Wege zur Informationsgewinnung entwickelt. Eine Möglichkeit sind so genannte Just-in-time Information Retrieval Agents (JITIR-Agents)[RM00]. Sie beobachten im Hintergrund den Kontext des Benutzers und versuchen aus den so erhaltenen Informationen eine Suchanfrage an eine Datenbank oder ein Recommender-System zu schicken. Die gewonnenen Informationen werden dann möglichst unaufdringlich dem Benutzer angezeigt. Er kann sich nun entscheiden diese Informationen genauer zu betrachten oder mit seiner Arbeit fortzufahren. Die kognitive Belastung bleibt hierbei sehr gering. JITIR-Agents reduzieren auf diese Weise enorm den Aufwand Informationen zu finden[RM00]. Durch ihre Funktionsweise sind sie jedoch nicht so exakt wie klassische Suchmaschinen, da sie nur 'erraten' was den Benutzer gerade interessieren könnte. Wenn ein Benutzer einer genaue Vorstellung von der Suchanfrage oder den Ergebnissen hat haben klassische Suchmaschinen Vorteile gegenüber den JITIR-Agents [RM00]. Ein Beispiel für ein solches System ist das EEXCESS Chrome Plugin. Es analysiert die aktuell besuchten Seiten des Nutzers und schlägt ihm auf Grund der erlangten Daten am Rand des Browser-Fensters weiterführende Quellen aus der Europeana-Datenbank<sup>3</sup> vor. Ziel dieser Bachelorarbeit ist es, auf Basis des EEXCESS Plugins ein Chrome Plugin zu entwerfen und zu implementieren, welche eine Webseite in einzelne Paragraphen aufteilt und zu jedem Paragraphen kulturelle Inhalte der Europeana-Datenbank vorschlägt. Dabei soll das Design des Plugins helfen, die bei der in Kapitel 5 erläuterten Evaluierung aufgetauchten Probleme zu beheben und Schwachstellen zu verbessern.

---

<sup>1</sup><https://www.google.com/>

<sup>2</sup><http://www.baidu.com/>

<sup>3</sup><http://www.europeana.eu/portal/>



## 1.1 Unterschiede Web Augmentation, Web Personalization, Web Customization

WA is to the web what augmented reality is to the physical world

## 1.2 Vormarsch von JITIR/Web Augmentation

JITIR = Just-in-Time Information Retrieval

- Möglichkeit das Web mit Funktionen anzureichern
- JITIR hat Vorteile/Nachteile gegenüber klassischen Suchmaschinen
  - Vorteil Suchmaschinen: Wenn Benutzer klare Vorstellung von der Suche/ Suchanfrage hat oder genau weiß wonach er sucht
  - Vorteil JITIR: Aktueller Task muss nicht komplett unterbrochen werden. Man verliert nicht den Überblick was man gerade macht/bleibt im Kurzzeitgedächtnis. JITIR Agents greatly reduce the cost of searching for information
  - Nachteil JITIR: Ergebnisse sind nicht so exakt wie bei Suchmaschinen

## 1.3 Verbesserung des EEXCESS Plugins / Unterschiede zum EEXCESS Plugin

Auswertung der alten Evaluierung? Erleichtern von Recherchen durch einbinden von weiterführenden Links zu kulturellen Inhalten direkt in die betrachtete Webseite

## 2 Related Work

### 2.1 Unterschiede und Gemeinsamkeiten zu JITIR-Agents

#### 2.1.1 Proaktiv vs. User Interaction

#### 2.1.2 Informationen darstellen in „Nonintrusive Manner“

#### 2.1.3 Awareness of user's local context

### 2.2 Vergleich mit existierenden JITIR Agents

#### 2.2.1 Remembrance Agent (in EMACS Editor)

Zeigt Quellen an auf Basis des geschriebenen Texts, Benutzer kann Suchanfrage dann auch noch manuell anpassen/verfeinern -  $\checkmark$  Vorteil von Suchmaschinen wird mit integriert

#### 2.2.2 Margin? Web Plugin ähnlich wie EEXCESS

#### 2.2.3 Watson

#### 2.2.4 EEXCESS?

### 2.3 Text Retrieval Algorithmen

Term Frequency/Inverse Document Frequency algorithm Text rank

### 2.4 Unterschiede und Gemeinsamkeiten zu „Automatic help systems“ (z.B. Microsoft Office Assistant - $\checkmark$ Domain spezifisch)

Domain spezifisch vs. Domain unabhängig

## 3 Konzept

### 3.1 Warum kein Proactiver JITIR-Agent?

-  $\checkmark$  API Limitierung und decrease cognitive load -  $\checkmark$  Benutzer entscheidet ob er weitere Informationen erhalten möchte -  $\checkmark$  Benutzer kann Suchanfrage erst anpassen (Nachteil von Margin Notes (Paper 4))

### **3.2 Anzeige der Ergebnisse**

### **3.3 Erklärung der Such-Anfragen Generierung**

### **3.4 Anpassen der Suchanfrage durch den Nutzer**

### **3.5 Verbesserung der Suchanfrage z.B. durch maschinelles Lernen**

## **4 Implementierung**

### **4.1 Verwendung von AngularJS für alle Komponenten des Plugins**

### **4.2 Bau der GUIs**

- ⚡ Darf den Benutzer nicht zu sehr ablenken - ⚡ Ergebnisse müssen in der Nähe ihrer „Quelle“ angezeigt werden (proximity compatibility principle) - ⚡ Benutzer muss klar zwischen Webseite und Augmentation unterscheiden können - ⚡ bunt, auffälliges Design - ⚡ Ramping interface: Mehr Benutzerinteraktionen führen zu mehr angezeigten Informationen (Erklärung der Stages)

### **4.3 Einbindung der REST-Services**

## **5 Evaluierung?**

Aufgaben die Benutzer mit EEXCESS Lösen mussten müssen sie jetzt mit Redesign lösen.  
Vergleich der Ergebnisse?

## **6 Future Work**

### **6.1 Alternative Algorithmen zur Textanalyse**

### **6.2 Implementierung einer automatischen**

#### **Suchanfragen-Verbesserung durch maschinelles Lernen**

- Mehr kontextuelle Informationen miteinbeziehen
- Such-Profil des Nutzers erstellen

### **6.3 Verbesserung der Ergebniss-Güte**

- durch Query Expansion
- durch Filtern der Ergebnisse (mehr Präzision da Ausbeute bei JITIR nicht so relevant)

## **6.4 Anpassung der Anwendung auf mobile Nutzung**

## **7 Conclusion**

- Steigerung der Effektivität und Produktivität von wissenschaftlichem Arbeiten
- Starke Effektivitätssteigerung durch Punkte aus Future Work möglich?

## 8 Quellenverzeichnis

- [AI] ALEXA INTERNET, Inc. 1996 2.: *Alexa*. <http://www.alexa.com/topsites>. – Zugriff: 13.05.2015, Archiviert mit WebCite®: <http://www.webcitation.org/5hgZUZacN>
- [RM00] RHODES, Bradley J. ; MAES, Pattie: Just-in-time information retrieval agents. In: *IBM Systems journal* 39 (2000), Nr. 3.4, S. 685–704

## Anhang

### GUI Screenshots

Unterkategorie, die nicht im Inhaltsverzeichnis auftaucht.

# Erklärung

Hiermit versichere ich, dass ich meine Abschlussarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Datum:

.....

(Unterschrift)